

Zwei-Phasen-Pretesting

Prüfer, Peter; Rexroth, Margrit

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Prüfer, P., & Rexroth, M. (2000). *Zwei-Phasen-Pretesting*. (ZUMA-Arbeitsbericht, 2000/08). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-200864>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

ZUMA-Arbeitsbericht 2000/08
Zwei – Phasen – Pretesting
Peter Prüfer und Margrit Rexroth
August 2000
ISSN 1437-4110

ZUMA
Quadrat B2,1
Postfach 12 21 55
68072 Mannheim
Telefon: 0621-1246-227/230
Telefax: 0621-1246-100
E-mail: pruefer@zuma-mannheim.de
rexroth@zuma-mannheim.de

Zwei – Phasen – Pretesting ¹

Zusammenfassung:

Einen Fragebogen vor Beginn der Hauptstudie einem Test zu unterziehen, ist ein allgemein anerkannter methodischer Standard bei der Durchführung sozialwissenschaftlicher Umfragen. Dabei werden neben dem klassischen Testverfahren, dem Standard-Pretest, seit mehr als 10 Jahren auch sogenannte kognitive Techniken eingesetzt, die sich vor allem zur Überprüfung des Fragenverständnisses einzelner Fragen eignen. Im vorliegenden Bericht wird als Hilfestellung für die Praxis ein Verfahren – das „Zwei-Phasen-Pretesting“ - vorgestellt, das in zwei Schritten sowohl die Anwendung kognitiver Techniken als auch den Einsatz des Standard-Pretests vorsieht. Die Besonderheit des Verfahrens besteht in einer von den Autoren entwickelten standardisierten, von Interviewern im Feld anwendbaren Form kognitiver Techniken, dem Evaluationsinterview, das im ersten Schritt zum Einsatz kommt.

1. Einleitung

Die Durchführung und Auswertung von Pretests zur Verbesserung der Qualität von Fragebogen ist eine der Leistungen, die ZUMA für externe Forscher erbringt. Lange Zeit war der sogenannte Standard-Pretest² das einzige dabei angewandte Verfahren. Dabei wird der Fragebogen nach Abschluss der Fragebogenentwicklung im Feld unter Einsatz von Interviewern getestet. Die Situation änderte sich durch die neuen kognitiven Pretest-Techniken³, die es ermöglichten, gezielt und systematisch Informationen speziell über das Fragenverständnis seitens der Befragten zu sammeln, eine Fähigkeit, die der Standard-Pretest nicht besitzt. Andererseits ist der Standard-Pretest in der Lage, über wichtige Qualitätsmerkmale des Fragebogens, wie z.B. die Handhabung durch Interviewer Auskunft zu geben, was wiederum kognitive Techniken nicht leisten können. Die daraus resultierende Lösung für ein neues Pretestverfahren konnte also nicht heißen „kognitive Techniken oder Standard-Pretest“, sondern nur „kognitive Techniken und Standard-Pretest“. Unter diesen Voraussetzungen galt es, ein Verfahren zu entwickeln, das beide Techniken integriert und

¹ Eine verkürzte Version dieses Berichts ist erschienen in: Mohler/Lüttinger (2000).

² Vgl. z.B. Friedrichs (1973); Karmasin/Karmasin (1977); Sudman/Bradburn (1982).

³ Vgl. z.B. Jabine/Straf/Tanur/Tourangeau (1984), Forsyth/Lessler (1991), Schwarz/Sudman (1996)

darüber hinaus den besonderen Bedingungen gerecht wird, unter denen bei ZUMA Pretests durchgeführt werden⁴.

Das Ergebnis ist das von den Autoren entwickelte Zwei-Phasen-Pretesting. Die Besonderheit dieses Verfahrens besteht darin, dass in der ersten Phase einzelne Fragen mittels eines Evaluationsinterviews⁵, d.h. durch standardisierte Anwendung kognitiver Techniken von Interviewern im Feld getestet werden; in der zweiten Phase wird dann der gesamte Fragebogen mittels eines Standard-Pretests überprüft. Im Folgenden wird das Verfahren detailliert beschrieben, wobei zunächst die methodischen Grundlagen im Vordergrund stehen.

1.1 Was ist eine gute Frage?

Noch vor etwa 30 Jahren galt die Konstruktion von Fragen und Fragebogen als mehr oder weniger reine Kunstlehre⁶. Mittlerweile existiert eine Vielzahl von empirischen Befunden und Erfahrungen, die eine Hilfestellung für die Konstruktion guter Fragen geben können⁷. Allerdings gibt es bis heute weder eine empirisch fundierte, umfassende Theorie der Frage noch ein komplettes und hinreichend konkretes Regelwerk, mit dessen Hilfe allein es möglich wäre, gute Fragen zu konstruieren.

Doch was ist eine gute Frage?

Eine gute Frage führt zu einer Antwort, die das, was wir beschreiben oder erklären wollen, reliabel und valide mißt⁸.

Diese Definition einer guten Frage ist sehr abstrakt und theoretisch formuliert und eignet sich deshalb eher als generelle Zielvorgabe und weniger als konkrete Hilfe in der Praxis, wenn es darum geht, gute Fragen zu konstruieren.

1.2 Wie konstruiert man gute Fragen?

Ein sinnvoller erster Schritt auf dem Weg zu einer guten Frage besteht u. E. darin, sich bei der Konstruktion zunächst an grundlegenden Eigenschaften bzw. Qualitäten zu orientieren, die eine Frage besitzen muss, um reliable und valide Antworten zu produzieren⁹. Fowler (1995) nennt als wichtigste Eigenschaften fünf Grundqualitäten:

- **Eine Frage muss so konstruiert sein, dass sie von allen Befragten konsistent im Sinne des Forschers verstanden bzw. interpretiert wird.**

⁴ Auf diese besonderen Bedingungen wird später noch detailliert eingegangen.

⁵ Das Evaluationsinterview wird in Kapitel 2.3 näher beschrieben.

⁶ Vgl. z.B. Payne (1951).

⁷ Vgl. z.B. Hippler/Schwarz/Sudman (1987); Jobe/Loftus (1991); Sudman/Bradburn/Schwarz (1996)

⁸ Diese Definition lehnt sich stark an Fowler (1995) an.

⁹ Wir beziehen uns in diesem Bericht ausschließlich auf sog. Survey-Interviews.

- **Eine Frage muss so konstruiert sein, dass sie in konsistenter Weise erhoben werden kann.**

Dies ist immer dann nicht der Fall, wenn die Vorgehensweise für den Interviewer nicht eindeutig definiert ist (z.B. fehlende oder missverständlich formulierte Anweisungen, Texte in Klammern).

- **Eine Frage muss so konstruiert sein, dass sie den Befragten in konsistenter Weise vermittelt, welche Art bzw. Form von Antwort verlangt wird.**

Beispiel für eine Frage, die Art bzw. Form der Antwort offen lässt:

„Wann sind Sie nach Mannheim gezogen?“

Mögliche Formen der Antwort:

„Mit 18 Jahren“

„Im Jahr 1975“.

„Nach Beendigung meines Studiums“

Damit die erwünschte Form der Antwort für alle Befragten eindeutig ist, müsste die Frage lauten: „In welchem Jahr sind Sie nach Mannheim gezogen?“

- **Befragte müssen die zur Beantwortung der Frage nötigen Informationen in ausreichendem Maß besitzen (Ausnahme: Es soll tatsächlich Wissen abgefragt werden).**
- **Befragte müssen bereit sein, korrekte und valide Antworten zu geben.**

Die Qualität von Daten wird seitens des Erhebungsinstruments in der Praxis jedoch nicht nur von Eigenschaften der einzelnen Frage beeinflusst, sondern darüber hinaus auch von Eigenschaften des gesamten Fragebogens bzw. von Faktoren, die aus dem Zusammenwirken mehrerer Fragen entstehen (können).

Als weitere wichtige Eigenschaftsziele wären hier zu nennen:

- **Ein Fragebogen muss so konstruiert sein, dass sowohl die Motivation, alle Fragen korrekt zu beantworten, als auch die Konzentration der Befragten in keiner Weise beeinträchtigt wird.**

Die beiden wichtigsten Faktoren, die diese Eigenschaft beeinflussen können, sind das Thema/die Themen der Befragung (z.B. Aktualität, Relevanz/Bedeutung für die Befragten oder Kompetenz seitens der Befragten) sowie die Dauer der Befragung (mit zunehmender Befragungsdauer steigt die Gefahr abnehmender Konzentration und Motivation seitens der Befragten).

- **Ein Fragebogen muss so konstruiert sein, dass von einzelnen Fragen ausgehende unerwünschte Einflüsse/Effekte auf nachfolgende Fragen ausgeschlossen werden (Kontexteffekte/Sukzessionseffekte).**

- **Ein Fragebogen muss so konstruiert sein, dass Interviewer keine Probleme mit dem Ablauf der Erhebung des gesamten Fragebogens haben.**

Interviewer müssen z.B. problemlos feststellen können, welche Fragen in welcher Reihenfolge gestellt werden müssen und welche Fragen nicht zu stellen sind. Probleme können hier besonders bei einer komplizierten Filterführung auftreten.

Wie kann man feststellen, ob bzw. in welchem Umfang eine Frage bzw. ein Fragebogen die aufgeführten Eigenschaften besitzt?

Eine Möglichkeit, dies zu überprüfen wäre, einen „Check“ durchzuführen, der sich ausschließlich an der Erfahrung des Fragebogenkonstruktors orientiert. Nicht ohne Grund warnen führende Fragebogenexperten vor dieser Vorgehensweise¹⁰ und empfehlen stattdessen, vor der endgültigen Erhebung möglichst intensive und systematische empirische Evaluationsmaßnahmen durchzuführen.

Allerdings kann man selbst nach einer noch so intensiven Evaluation niemals völlig sicher sein, dass eine Frage bzw. ein Fragebogen auch Daten von perfekter Qualität im Sinne der genannten Definition generieren wird, aber man kann damit das Risiko, „schlechte“ Fragen zu stellen oder den Mangel an wichtigen Eigenschaften des Fragebogens zu übersehen, erheblich reduzieren.

1.3 Das Zwei-Phasen-Pretesting-Konzept

Wie bereits erwähnt, wurden bis zum Bekanntwerden kognitiver Pretest-Techniken bei ZUMA – wie in der gesamten Profession üblich – ausschließlich Standard-Pretests durchgeführt. Als wir begannen, die „neuen“ kognitiven Techniken in die Pretestpraxis umzusetzen, entwickelten wir bei ZUMA unter dem Begriff „Multi-Method-Pretesting“ ein Konzept, das für den Einsatz unterschiedlicher Testverfahren im Verlauf der Fragebogenentwicklung steht.¹¹ Es handelt sich dabei um ein Konzept, das vor der Durchführung eines Standard-Pretests ganz generell die Anwendung kognitiver Techniken¹² vorschreibt, konkrete Modalitäten der Anwendung aber nicht anspricht. Unter diesem allgemeinen Rahmen stellt das „Zwei-Phasen-Pretesting“ eine Variante dar, die den Einsatz kognitiver Techniken auf konkrete Art und Weise angeht, wobei es nicht zuletzt aus der Notwendigkeit heraus entstand, den Einsatz kognitiver Techniken den besonderen Bedingungen anzupassen, unter denen bei ZUMA Pretests durchgeführt werden.

¹⁰ Vgl. z.B. Sudman/Bradburn (1982).

¹¹ Siehe hierzu Prüfer/Rexroth (1996)

¹² Hierbei sind auch Verfahren wie Focus Groups mit eingeschlossen.

Ein wesentliches Merkmal dieser besonderen Bedingungen besteht darin, dass die zu testenden Fragen in der Regel bereits in Form eines vom externen Forscher relativ weit entwickelten und erhebungsfähigen Fragebogens vorliegen. Dies bedeutet, dass Evaluationsverfahren- und -techniken, die speziell in der frühen Entwicklungsphase eines Fragebogens zur Anwendung kommen (z.B. Focus Groups oder der Einsatz kognitiver Techniken im Labor mit wenigen, willkürlich ausgewählten Befragten), in dieser fortgeschrittenen Phase als Evaluationsinstrumente nicht mehr in Betracht kommen.

Ein weiteres Merkmal der besonderen ZUMA-Bedingungen ist der vorherrschende Wunsch der externen Forscher, auch bei der Anwendung kognitiver Techniken mit systematisch ausgewählten Stichproben mit einem Minimum von ca. 20 Fällen zu arbeiten.

Ein drittes Merkmal schließlich ist in der Regel der Mangel an zeitlichen und finanziellen Ressourcen seitens der externen Forscher.

Unter Berücksichtigung sowohl dieser besonderen Bedingungen als auch der methodischen Zielvorgaben bzw. Qualitätsstandards haben wir das Zwei-Phasen-Pretesting entwickelt, das keinesfalls eine maximale Evaluation zum Ziel hat, sondern lediglich in der Lage sein soll, in ausreichendem Maße Informationen über die Existenz oder das Fehlen zentraler Qualitätsmerkmale sowohl bezüglich einzelner Fragen als auch des gesamten Fragebogens zu liefern.

Das Verfahren arbeitet in zwei Schritten¹³:

- **Phase 1: Anwendung kognitiver Techniken in Form standardisierter Evaluationsinterviews durch Interviewer im Feld**
- **Phase 2: Durchführung eines Standard-Pretests**

Unser Bericht beschränkt sich im Wesentlichen auf die Beschreibung der praktischen Anwendung des Zwei-Phasen-Pretestings, wobei unsere besondere Aufmerksamkeit den kognitiven Techniken gilt.

¹³ Bei ZUMA geht diesen beiden Schritten noch eine Expert Review (Expertenbegutachtung) voraus.

Im folgenden Kapitel wird zunächst die Vorgehensweise in Phase 1 beschrieben.

2. Phase 1: Anwendung kognitiver Techniken mittels Evaluationsinterviews

2.1 Zu kognitiven Techniken allgemein

Im Pretestbereich werden kognitive Techniken vor allem zur Überprüfung des Fragenverständnisses eingesetzt, d.h. es wird mit ihrer Hilfe getestet, wie Befragte Begriffe und Formulierungen verstehen, auf Grund welcher Überlegungen Antworten zustande kommen. Man bezeichnet diese Techniken als aktive Techniken, denn im Gegensatz zum Standard-Pretest, bei dem Probleme nur dann explizit werden, wenn Befragte diese von sich aus formulieren, werden beim Einsatz kognitiver Techniken Befragte – unabhängig von ihren Antwortreaktionen - gezielt zu Erklärungen, Begründungen oder einer Offenlegung von Gedankengängen aufgefordert. Dadurch sind solche Techniken gut geeignet, Verständnisprobleme bei Fragen aufzudecken.

Um welche Techniken geht es dabei im einzelnen? Tatsächlich gibt es eine Vielzahl unterschiedlicher kognitiver Techniken¹⁴; an dieser Stelle sollen jedoch lediglich die Techniken erwähnt werden, die beim Testen von Survey-Fragen am häufigsten zum Einsatz kommen.

Wichtige Kognitive Techniken:

- **Probing:** Hinterfragen von Antworten

Tabelle 1: Probing-Varianten

Variante	Beschreibung
Special Comprehension Probing	Der/die Befragte soll bestimmte Aspekte oder Begriffe einer Frage erklären.
Information Retrieval Probing	Der/die Befragte soll den Vorgang der Informationsbeschaffung beschreiben.
Category Selection Probing	Der/die Befragte soll begründen, warum er/sie sich für bestimmte Antwortvorgaben oder Skalenwerte entschieden hat.
General Probing	Dem/der Befragten wird eine sehr allgemeingehaltene Zusatzfrage, z.B. nach dem Verständnis gestellt („Gibt es etwas, was Sie nicht verstanden haben?“)

¹⁴ Einen deutschsprachigen Überblick zu kognitiven Pretestverfahren geben Prüfer/Rexroth (1996).

- **Think Aloud:** Methode des lauten Denkens

Tabelle 2: Varianten des Think Aloud

Variante	Beschreibung
Concurrent Think Aloud	Der/die Befragte wird aufgefordert, während der Beantwortung der Frage laut zu denken, d.h., alle Gedankengänge, die zur Antwort führen, laut zu formulieren.
Retrospective Think Aloud	Nach Beantwortung der Frage wird die Zielperson gefragt, auf Grund welcher Überlegungen ihre Antwort zustande kam.

- **Paraphrasing:** Der/die Befragte wird nach Beantwortung der Frage aufgefordert, den Fragetext in eigenen Worten zu wiederholen.
- **Confidence Rating:** Der/die Befragte bewertet nach der eigentlichen Antwort den Grad der Verlässlichkeit dieser Antwort (meist mittels einer Skala).

Dass kognitive Techniken in der Lage sind, Formierungsdefizite bei Fragen verlässlicher aufzudecken als dies der Standard-Pretest kann, ist mittlerweile hinreichend bekannt¹⁵. Anhand eines Beispiels soll die Leistungsfähigkeit kognitiver Techniken demonstriert werden. Wir beziehen uns dabei auf die Technik eines „category selection probings“, das im Rahmen eines kognitiven Pretests bei 6 Zielpersonen in der Laborsituation zur Überprüfung des inhaltlichen Verständnisses eingesetzt wurde. Ziel der Verständnisüberprüfung war unter anderem eine für eine Umfrage ungewöhnlich schwierige politische Aussage, die mittels einer 7er-Skala zu bewerten war.

Der Text der Aussage lautete:

„Das Ziel der Nationalsozialisten war, durch den Krieg gegen die Sowjetunion den jüdischen Bolschewismus zu beseitigen.“

(Skalenwert 1 = stimme überhaupt nicht zu; Skalenwert 7 = stimme voll und ganz zu.)

Unmittelbar nach der Bewertung des Items wurde folgende Probingfrage gestellt:

„Und warum haben Sie sich für diesen Skalenwert entschieden?“

¹⁵ Siehe hierzu besonders Presser/Blair (1994).

Ergebnis:

Die folgende Auflistung zeigt fallspezifisch den genannten Skalenwert und die Antwort auf die Probingfrage.

Tabelle 3: Antworten auf Probingfragen

N = 6

Fall	Skalenwert	Antwort auf Probingfrage (wörtlich)
1	4	„Das weiß ich nicht, ich kann mit der Aussage nichts anfangen.“
2	2	„Ich weiß es nicht, aber ich glaube es nicht. Habe die 2 vergeben auf Grund meines Bauchgefühls.“
3	4	„Bin mir relativ unsicher. Ich dachte, ich nehme einen Mittelweg. Die 4 deshalb, weil ich es nicht weiß.“
4	1	„Weil ich mir einfach nicht vorstellen kann, dass die Nationalsozialisten ein anderes Ziel hatten, als in dem Fall Land zu gewinnen. Vermute es aber, habe keinen blassen Schimmer.“
5	6	„Ich vergebe ungern 1 und 7. Mit der 6 möchte ich relativieren und ich glaube, dass es das Ziel der Nationalsozialisten war.“
6	5	„Es war immer das Ziel der Nationalsozialisten gegen die Juden zu handeln.“

In allen 6 Fällen wurden Skalenwerte genannt. Geht man davon aus, dass die Probingfrage nicht gestellt worden wäre, so gäbe es keinerlei Hinweise auf Verständnisprobleme. Erst die aktive Nachfrage ergab, dass in vier der sechs Fälle die genannten Werte nicht valide waren, da die Befragten zugaben, von den Inhalten eigentlich gar nichts zu verstehen und nur auf Vermutung hin oder weil es von ihnen gefordert wurde, einen Wert nannten. Die Tatsache, dass Zielpersonen aus Gründen der Unsicherheit nicht nur auf die Mittelkategorie der Skala als eine Art „weiß nicht“ Kategorie ausweichen (Fall 3), sondern sich auch inhaltlich bedeutungsvollen Extremwerten wie „1“ und „2“ zuordnen (Fall 4 und Fall 2), zeigt, wie wichtig es ist, Fragen im Vorfeld einer Untersuchung auf ihr Verständnis zu überprüfen.

Es soll hier nicht weiter auf die Funktionstüchtigkeit kognitiver Techniken eingegangen werden. Wir wollen uns nun der praktischen Anwendung dieser Techniken zuwenden und zeigen, wie diese in die Erhebungspraxis integriert werden können, ein bis jetzt in der Literatur kaum dokumentierter Aspekt. Im folgenden werden mögliche Anwendungsformen vorgestellt.

2.2 Anwendungsmodalitäten beim Einsatz kognitiver Techniken

2.2.1 Traditionelle Vorgehensweise

Kognitive Techniken stammen ursprünglich aus dem Arbeitsbereich von Kognitionspsychologen. Als Sozialwissenschaftler begannen, diese Techniken zur Erforschung des Frage-Antwort-Prozesses zu nutzen, haben sie auch weitgehend die traditionelle Vorgehensweise übernommen. So werden diese Techniken auch in sozialwissenschaftlichen Studien in der Regel unter sog. Laborbedingungen von Experten – häufig sind das die Wissenschaftler oder Mitarbeiter des Projektteams selbst – bei nur wenigen, d.h. bis maximal 10 Testpersonen in einer weitgehend offenen Gesprächsform angewandt. Allerdings wird meist auf die klassische Ausstattung mit Videorecorder oder Einwegscheibe verzichtet. Zwingend notwendig ist jedoch die Aufzeichnung der Gespräche auf Tonband.

Charakteristisch für den traditionellen Einsatz kognitiver Techniken ist eine eher unstrukturierte Vorgehensweise. Es bleibt in das Belieben des Experten gestellt, wie er bei der Überprüfung einer Frage vorgeht, wie intensiv er z.B. die Reaktionen der Zielperson hinterfragt und welche Techniken er im einzelnen anwendet. Voraussetzung dabei ist, dass der Experte die Techniken in genauer Kenntnis der Zielsetzung der zu überprüfenden Frage anwendet. Wenn wir hier also von „Experte“ sprechen, so meinen wir eine Person, die mit dem theoretischen Hintergrund der Fragen vertraut ist.

Diese traditionelle Vorgehensweise hat unserer Ansicht nach Vor- und Nachteile:

Ein Vorteil ist, dass der Experte jede für das Fragenverständnis aufschlussreiche Befragtenreaktion auf Grund seines Hintergrundwissens als solche identifizieren kann. Ein weiterer Vorteil ist, dass er durch seine nur wenig an Strukturen gebundene, flexible Vorgehensweise die eigentlichen Ursachen für Befragtenprobleme detailliert aufdecken kann. Der Nachteil eines solchen Vorgehens liegt darin, dass es häufig zu eher unsystematischen Eindrücken führt, was die Qualität der zu überprüfenden Frage angeht. Die erhobenen Informationen können auf Grund der in den einzelnen Gesprächen unterschiedlich gesetzten Stimuli meist nicht direkt miteinander verglichen werden. Hinzu kommt die geringe Fallzahl und die zumeist unsystematische Auswahl von Befragungspersonen, die die Aussagekraft der Testergebnisse einschränken.

2.2.2 Standardisierte Vorgehensweise

Als einen Weg zur Systematisierung sehen wir die Entwicklung eines speziellen Evaluationsfragebogens, in den kognitive Techniken zur Überprüfung von Fragen

eingearbeitet werden, und der im Rahmen eines Test-Interviews standardisiert erhoben wird. Wir nennen diese Form des Interviews „Evaluationsinterview“.

Die streng standardisierte Erhebungsform hat zwar den Nachteil, dass vom Interviewer keine zusätzlichen spontanen Nachfragen gestellt werden können, sie hat jedoch andererseits den Vorteil, „normale“ Pretestinterviewer einsetzen zu können, da die kognitiven Techniken lediglich wie offene Fragen erhoben werden. Ein weiterer Vorteil beim Einsatz von Interviewern liegt in der Möglichkeit zur Realisierung einer größeren Zahl von Evaluationsinterviews in relativ kurzer Zeit. Auch die Auswertung der Interviews ist auf Grund der Standardisierung des Evaluationsfragebogens relativ schnell durchführbar.

Der Einsatz kognitiver Techniken mittels eines standardisierten Evaluationsfragebogens, der durch Interviewer erhoben wird, ist eine bislang nicht praktizierte Vorgehensweise. Sie wurde bei ZUMA entwickelt, nicht zuletzt auch in Anbetracht der spezifischen Anforderungen¹⁶, die den Einsatz von Interviewern hier notwendig machen. Sowohl die Quantität als auch die Qualität der durch Evaluationsinterviews erhobenen Informationen ist wesentlich von der Konstruktion des Evaluationsfragebogens abhängig, ein Arbeitsschritt, der sich in der Praxis als schwierig und zeitintensiv erwiesen hat und auf dessen Ablauf im folgenden kurz eingegangen werden soll.

2.3 Zur Konstruktion eines Evaluationsfragebogens

Unter einem Evaluationsfragebogen verstehen wir einen eigens für das Evaluationsinterview konstruierten Fragebogen. Er enthält sowohl die eigentlichen Fragen, die auf ihre Qualität hin überprüft werden sollen, als auch die entsprechenden kognitiven Techniken, mittels derer die Überprüfung vorgenommen werden soll.

Vor der Entscheidung für eine kognitive Technik steht die Qualitätsbestimmung der zu überprüfenden Frage. Hierfür bieten sich sogenannte "Expertenratings" an, bei denen idealerweise mehrere Experten auf Grund ihrer praktischen Erfahrungen und Methodenkenntnis im Erhebungsbereich die Fragen bewerten. Hilfreich für eine solche Bewertung ist auch die Verwendung von Kriterienkatalogen wie z.B. die eingangs dargestellte Eigenschaftsliste von Fowler¹⁷ oder ein von Lessler und Forsyth (1995) entwickeltes Codesystem zur Merkmals- und Eigenschaftsbestimmung von Fragen hinsichtlich ihres kognitiven Anforderungsprofils. Zusätzlich ist bei diesem Arbeitsschritt die

¹⁶ Siehe Kapitel 1.3

¹⁷ Eine ähnliche Liste findet sich auch bei Morton-Williams (1979).

Zusammenarbeit mit dem inhaltlichen Forscher notwendig, da das Frageziel, d.h. das, was mit einer Frage gemessen werden soll, eine wesentliche Rolle bei der Qualitätsbestimmung der Frage spielt.

Unter Berücksichtigung der vermuteten bzw. erkannten Konstruktionsschwäche der Frage wird im zweiten Schritt die kognitive Technik bestimmt, die am besten in der Lage ist, die benötigten Informationen zu liefern. Bei der Entscheidung für eine bestimmte kognitive Technik muss vor allem berücksichtigt werden, wie direkt bzw. gezielt auf das Problem zugegangen werden soll. Wenn z.B. das Verständnis oder die Interpretation eines speziellen Begriffes eruiert werden soll, so wird man die Technik des „Special Comprehension Probing“ einsetzen und eine gezielte Nachfassfrage formulieren (Beispiel aus dem kognitiven Pretest zum Allbus 98: „Was verstehen Sie unter einem Bürgerbegehren?“). Wenn man speziell die Bedeutung eines gewählten Skalenwertes erfahren möchte, so wählt man ein „Category Selection Probing“ als Nachfrage („Warum haben Sie sich für diesen Skalenwert entschieden?“).

Anders verhält es sich bei Fragemängeln, die man nicht gezielt angehen möchte bzw. kann. Hier sind Techniken wie „Think Aloud“ (Methode des lauten Denkens bei der Antwortfindung), „Paraphrasing“ (Wiederholung des Fragetextes in eigenen Worten) oder „Information Retrieval Probing“ (bei Rückerinnerungsfragen: „Wie sind Sie zu dieser Antwort gekommen?“) angebracht, die den Bedeutungsgehalt der zu überprüfenden Frage eher generell explorieren. Aber auch sogenannte „General Probes“ (z. B.: „Gibt es etwas, das Sie bei der Frage nicht verstanden haben?“) können in solchen Fällen zum Einsatz kommen. Allerdings führen Probingfragen dieser Art nicht immer zu brauchbaren Ergebnissen, d.h. es werden nur selten konkrete Verständnisprobleme aufgedeckt. Zum einen mag die geforderte Verbalisierung eines Verständnisproblems manchen Befragten intellektuell überfordern, zum anderen kann eine Testperson die Frage nach Verständnisproblemen nur dann bejahen, wenn sie sich über ihr Unverständnis im klaren ist. In vielen Fällen sind sich Befragte eines falschen Fragenverständnisses aber gar nicht bewusst. In solchen Fällen können unspezifisch formulierte Nachfassfragen auch keine Verständnisprobleme aufdecken.

Im Evaluationsfragebogen sind die Fragen so angeordnet, dass der Zielperson zuerst die zu überprüfende „Originalfrage“ gestellt wird und unmittelbar nach deren Beantwortung eine oder mehrere speziell konstruierte Fragen oder Verfahrensweisen der jeweiligen kognitiven Technik. Dieses Vorgehen hat den Vorteil, dass neben den Ergebnissen aus den kognitiven Techniken sowohl die Antwortverteilungen bei der eigentlichen Frage als zusätzlicher Indikator für die Qualität der Frage dienen kann als auch eventuell auftretende

Interviewerprobleme in Bezug auf die technische Handhabung der zu testenden Frage erkannt werden können.

2.4 Vorgehensweise beim Evaluationsinterview

Jedes Interview wird auf Tonband aufgezeichnet. Vor Beginn der Befragung werden die Befragten auf den besonderen Charakter des Interviews hingewiesen und mit ihrer damit verbundenen speziellen Rolle vertraut gemacht. Wesentlich dabei ist, den Befragten zu vermitteln, dass sie keine herkömmliche Befragtenrolle übernehmen, sondern als „Mithelfende“ einen wichtigen Beitrag zur Verbesserung des Fragebogens leisten sollen. Die Erfahrung zeigt, dass Befragte in Kenntnis dieser Aufgabenstellung sehr wohl bereit sind, auch schwierige oder heikle Testfragen zu beantworten oder relativ peinliche Prozeduren über sich ergehen zu lassen. Um die Motivation zur Mitarbeit an dieser recht schwierigen Aufgabe zu fördern, empfehlen wir, den Befragten ein Honorar zu zahlen.

2.5 Bewertung des Evaluationsinterviews

Die Erfahrungen, die bei ZUMA mit Evaluationsinterviews bisher gemacht wurden, sind überwiegend positiv und lassen sich folgendermaßen zusammenfassen:

- Entsprechend geschulte Interviewer sind für die Durchführung eines Evaluationsinterviews mit Erfolg einsetzbar.
- Durch den Einsatz von Interviewern kann eine relativ hohe Zahl von Evaluationsinterviews in relativ kurzer Zeit durchgeführt werden..
- Durch die standardisierte Form des Evaluationsinterviews wird eine Vergleichbarkeit über alle erhobenen Fälle möglich und führt dadurch zu einer höheren Zuverlässigkeit der Daten.
- Die Entwicklung eines Evaluationsfragebogens - insbesondere die Konzeption der Zusatzfragen für die kognitiven Techniken - ist zeitaufwendig und ohne genaue Kenntnis der inhaltlichen Zielsetzung der Fragen nicht effektiv durchführbar.

3. Phase 2: Der Standard-Pretest

3.1 Vorgehensweise beim Standard-Pretest

Im Rahmen unseres Zwei-Phasen-Pretesting-Konzepts ist der Standard-Pretest das zweite Verfahren, dem zur Überprüfung von Qualitätsmerkmalen bezüglich einzelner Fragen und des gesamten Fragebogens eine wesentliche Rolle zukommt.

Unter einem Standard-Pretest versteht man eine - in der Regel einmalige - Test-Erhebung eines Fragebogens nach Abschluss der Fragebogenkonstruktion.

Die Pretest-Interviews werden von Interviewern im Feld durchgeführt, und zwar so weit wie möglich unter den Rahmenbedingungen der eigentlichen Haupterhebung. Die Befragten sind nicht über den Testcharakter der Interviews informiert.

Ein zentrales Problem des Standard-Pretests liegt in seiner ausschließlich „passiven“ Strategie zur Identifizierung von Qualitätsdefiziten bei Fragen. Das zugrundeliegende Prinzip, ausschließlich aus der Reaktion bzw. Antwort der Befragten Rückschlüsse auf die Fragenqualität zu ziehen ist insofern problematisch, als dabei Probleme der Befragten bei der Beantwortung einer Frage auf Konstruktionsmängel eben dieser Frage zurückgeführt werden, während Fragen, bei denen Befragte formal „korrekt“ antworten, als „gut konstruiert“ oder zumindest „weitgehend fehlerfrei“ angesehen werden. Dass diese Annahme weitgehend nicht zutrifft, konnten Belson (1981 und 1986) und Schuman (1966) in ihren Studien nachweisen. Die Ergebnisse zeigen, dass Fragen trotz formal korrekter Antwort häufig falsch verstanden worden waren. Dieses Phänomen konnten wir beim Einsatz kognitiver Nachfasstechniken selbst oft beobachten.

Der Standard-Pretest erkennt ein Verständnisproblem bei einer Frage also nur dann, wenn dieses vom Befragten selbst angesprochen wird. Dieses Prinzip führt zu eher zufälligen Informationen.

Die Stärken des Standard-Pretests liegen damit weniger im Bereich der Überprüfung des Verständnisses von Fragen als vielmehr in der Identifizierung von Problemen, die mit dem Ablauf des gesamten Fragebogens unter realen Feldbedingungen zu tun haben. Dies können einerseits Probleme auf Seiten der Interviewer sein (z.B. technische Probleme mit Listen, Anweisungen, Vorgabekategorien oder der Filterführung), aber auch Probleme seitens der Befragten (z.B. Nachlassen der Konzentration, geringe Antwortbereitschaft). Und letztendlich ist der Standard-Pretest das einzige Verfahren, das relativ verlässlich die Zeitdauer einer Befragung überprüfen kann, da nur hier das Interview unter den Rahmenbedingungen der Hauptstudie stattfindet.

3.2 Vorgehensweise beim Standard-Pretest

Bei der Durchführung der Pretest-Interviews haben die Interviewer folgende Aufgaben zu erfüllen:

- **Durchführung des Interviews wie ein „normales“ Interview,**
d.h. streng nach den Regeln zur Durchführung standardisierter Interviews¹⁸. Darüber hinausgehende Aktivitäten der Interviewer, z.B. aktives Hinterfragen bei auftretenden Problemen sind nicht erlaubt.
- **Identifizierung von Problemen und Auffälligkeiten, die bei der Durchführung des Interviews auftreten.**

Häufig registrieren Interviewer Erhebungsprobleme auf der Basis subjektiver Wahrnehmung und - wie eine Studie von Kreiselmeier et al.¹⁹ belegte - auch sehr selektiv, was zu unzuverlässigen Pretesterkenntnissen führt. Um zu erreichen, dass Interviewer weitgehend objektiv und systematisch vorgehen, wurde bei ZUMA ein Schulungskonzept entwickelt, das den Interviewern eine spezifische Beobachtungsstrategie vorschreibt. Dabei beurteilen die Interviewer bei jeder Frage das Verhalten des Befragten direkt nach dem Vorlesen des Fragetextes danach, ob es (formal) adäquat ist oder nicht. Adäquat ist das Befragtenverhalten dann, wenn es sich um eine spontane und im Sinne der Fragestellung formal korrekte Antwort handelt. Alle anderen Verhaltensweisen des Befragten sind vom Interviewer als nicht adäquat einzustufen und müssen registriert werden. Die Aufgabe, systematisch jedes inadäquate Befragtenverhalten zu registrieren, entbindet den Interviewer von der Schwierigkeit der Entscheidung, welche Verhaltensweisen eines Befragten als problematisch zu werten sind und welche nicht. Somit ist das Erkennen von Problemen nicht mehr ausschließlich in das Belieben des Interviewers gestellt. Die oben beschriebene Verfahrensweise der Interviewer ist ein ZUMA-spezifisches Merkmal bei der Durchführung von Standard-Pretests.

- **Registrierung der Befragungsdauer**
(meist nicht nur die Gesamtdauer, sondern auch die Dauer von Fragebogensegmenten).

¹⁸ Vgl. dazu Prüfer, P. (1998): Anleitung zur Durchführung standardisierter Interviews. ZUMA-Technischer Bericht 98/02.

¹⁹ Vgl. dazu Kreiselmaier/Prüfer/Rexroth (1989)

- **Berichten aller bei der Durchführung der Interviews als nicht adäquat identifizierte Befragtenverhaltensweisen.**

Diese „Pretest-Reports“ oder auch „Erfahrungsberichte“ werden fragenspezifisch und separat für jedes Interview erstellt.²⁰

4. Zur Leistungsfähigkeit des Zwei-Phasen-Pretesting

Nach der ausführlichen Beschreibung stellt sich die Frage nach der Tauglichkeit des Zwei-Phasen-Pretestings. Wir möchten diese Frage abschließend klären, indem wir prüfen, inwieweit das Verfahren in der Lage ist, in ausreichendem Maße Informationen über die eingangs erwähnten grundlegenden Qualitätseigenschaften der einzelnen Frage sowie des gesamten Fragebogens zu liefern.

- **Eine Frage muss so konstruiert sein, dass sie von allen Befragten konsistent im Sinne des Forschers verstanden bzw. interpretiert wird.**

Es ist gerade die Stärke kognitiver Techniken (Phase 1), in systematischer Weise Informationen über das Fragenverständnis zu liefern. Die Evaluation dieser Eigenschaft ist mit dem Standard-Pretest (Phase 2) nur unsystematisch bzw. zufällig möglich (z.B. nur dann, wenn Befragte von sich aus Unverständnis äußern).

- **Eine Frage muss so konstruiert sein, dass sie in konsistenter Weise erhoben werden kann.**

Diese Eigenschaft ist vor allem mit der Aufgabenstellung der Interviewer verbunden. Da in beiden Phasen die eigentliche Frage in gleicher Form erhoben wird, können geschulte Interviewer – sowohl in Phase 1 als auch in Phase 2 - in weitgehend systematischer Weise Informationen darüber liefern, ob eine Frage diese Bedingungen erfüllt.

- **Eine Frage muss so konstruiert sein, dass sie den Befragten in konsistenter Weise vermittelt, welche Art bzw. Form von Antwort verlangt wird.**

In Phase 1 geben kognitive Verfahren die Möglichkeit, durch gezielte Nachfragen systematisch Informationen darüber zu sammeln, welche Form der Antwort am ehesten valide und reliable Daten liefern kann. Dagegen kann der Standard-Pretest (Phase 2) nur unsystematisch bzw. zufällig Hinweise liefern (z.B. dann, wenn die Antwort der Befragten nicht der intendierten Form entspricht).

²⁰ Es empfiehlt sich, zusätzlich zu den schriftliche Berichten noch ein sog. Interviewer-Debriefing, d.h. eine mündliche Befragung der Interviewer nach der Durchführung aller Interviews durchzuführen.

- **Befragte müssen die zur Beantwortung der Frage nötigen Informationen in ausreichendem Maß besitzen (Ausnahme: Es soll tatsächlich Wissen abgefragt werden).**

In Phase 1 besteht hier die Möglichkeit, durch Nachfragen zu evaluieren, in welchem Umfang Befragte die nötigen Informationen besitzen oder wie schwierig es ist, diese Informationen abzurufen. Hinweise über diesen Informationsbesitz kann der Standard-Pretest dagegen nur unsystematisch bzw. zufällig liefern.

- **Befragte müssen bereit sein, korrekte und valide Antworten zu geben.**

In Phase 1 kann diese Bereitschaft explizit abgefragt bzw. hinterfragt werden. Im Standard-Pretest (Phase 2) sind Informationen darüber nur dann erhältlich, wenn Befragte von sich aus Reaktionen äußern, die auf das Ausmaß dieser Bereitschaft schließen lassen (z.B. Verweigern der Auskunft).

- **Ein Fragebogen muss so konstruiert sein, dass die Motivation, alle Fragen korrekt zu beantworten, als auch die Konzentration der Befragten nicht beeinträchtigt wird.**

In Phase 1 können darüber so gut wie keine Informationen gesammelt werden, da einerseits der Fragebogen hier nicht in seiner eigentlichen Form erhoben wird (auf die einzelne Frage folgen jeweils kognitive Techniken) und andererseits die Aufgabenstellung der Befragten nicht derjenigen der eigentlichen Erhebung entspricht (Befragte sind hier eher Helfer bei der Fragenkonstruktion). Dagegen kann der Standard-Pretest in Phase 2 sowohl in unsystematischer Weise (durch spontane Äußerungen des Befragten) als auch in systematischer Weise (durch ein im Anschluss an das Interview durchgeführtes Befragten-Debriefing) Informationen über motivationsrelevante Aspekte liefern.

- **Ein Fragebogen muss so konstruiert sein, dass unerwünschte Einflüsse/Effekte von einzelnen Fragen auf nachfolgende Fragen ausgeschlossen werden (Kontexteffekte/Sukzessionseffekte).**

Durch die Anwendung kognitiver Techniken in Phase 1 können Informationen über die Existenz dieser unerwünschten Effekte gesammelt werden, allerdings nur dann auf systematische Weise, wenn durch gezieltes Probing geklärt werden kann, ob ein vermuteter Effekt existiert oder nicht. Beim Standard-Pretest können Hinweise auf die Existenz dieser Effekte wiederum nur unsystematisch gesammelt werden.

- **Ein Fragebogen muss so konstruiert sein, dass Interviewer keine Probleme mit dem Ablauf der Erhebung des gesamten Fragebogens haben.**

Der problemlose Ablauf der Erhebung des gesamten Fragebogens kann nur mit dem Standard-Pretest in weitgehend systematischer Weise überprüft werden. Kognitive Techniken sind dazu nicht in der Lage, da hier der Fragebogen nicht in seiner eigentlichen Form erhoben wird.

Die Leistungsfähigkeit des Zwei-Phasen-Pretestings bezüglich der Überprüfung der Qualitätseigenschaften einer Frage/eines Fragebogens ist zur besseren Übersicht abschließend noch einmal in tabellarischer Form dargestellt.

Tabelle 4: Leistungsfähigkeit des Zwei-Phasen-Pretesting:

++	Das Verfahren kann systematisch Informationen liefern.
+	Das Verfahren kann nur unsystematisch Informationen liefern.
-	Das Verfahren kann keine Informationen liefern.

Zu prüfende Eigenschaften	PHASE 1 Evaluations- Interview	PHASE 2 Standard- Pretest
Eine Frage muss so konstruiert sein, dass sie von allen Befragten im Sinne des Forschers verstanden/interpretiert wird.	++	+
Eine Frage muss so konstruiert sein, dass sie in konsistenter Weise erhoben werden kann.	++	++
Eine Frage muss so konstruiert sein, dass sie den Befragten in konsistenter Weise vermittelt, welche Art bzw. Form von Antwort verlangt wird.	++	+
Befragte müssen die zur Beantwortung der Frage nötigen Informationen in ausreichendem Maß besitzen (Ausnahme: Es soll tatsächlich Wissen abgefragt werden).	++	+
Befragte müssen bereit sein, korrekte und valide Antworten zu geben.	++	+
Ein Fragebogen muss so konstruiert sein, dass sowohl die Motivation als auch die Konzentration der Befragten, alle Fragen korrekt zu beantworten, in keiner Weise beeinträchtigt wird.	-	+
Ein Fragebogen muss so konstruiert sein, dass unerwünschte Einflüsse/Effekte von einzelnen Fragen auf nachfolgende Fragen ausgeschlossen werden (Kontexteffekte/Sukzessionseffekte).	++	+
Ein Fragebogen muss so konstruiert sein, dass Interviewer keine Probleme mit dem Ablauf der Erhebung des gesamten Fragebogens haben.	-	++

Der Tabelle ist zu entnehmen, dass die genannten Qualitätsaspekte einer Frage bzw. eines Fragebogens überwiegend durch das Evaluationsinterview bzw. durch die kognitiven Techniken in systematischer Weise überprüft werden können. Der Standard-Pretest zeigt sich als Verfahren nur bei solchen Aspekten überlegen, die im Zusammenhang mit dem Fragebogen als Ganzes stehen. In der Tabelle sind das die Bereiche „Befragtenmotivation/-konzentration“ (beeinflusst durch Faktoren wie z.B. Relevanz des Themas/von Einzelinhalten, Kompetenz seitens der Befragten, subjektiv empfundene Interviewdauer) und „Ablauf der Erhebung insgesamt“ (z.B. Filterführung).

Neben der Evaluation von Qualitätseigenschaften ist die Befragungsdauer ein weiteres relevantes Merkmal, das im Pretest überprüft werden muss, mit dem Ziel, eine möglichst genaue Schätzung der Interviewdauer in der Hauptstudie abgeben zu können. Dies kann nur der Standard-Pretest leisten, da hier der Fragebogen ausschließlich unter Hauptfeldbedingungen zum Einsatz kommt.

Als Fazit kann festgehalten werden, dass beide Elemente des Zwei-Phasen-Pretestings – Evaluationsinterview und Standard-Pretest – ihre Stärken und Schwächen haben und in gleichem Maße erforderlich sind, um über die zentralen Qualitätsaspekte einer Frage bzw. eines Fragebogens – mehr oder weniger verlässlich - Auskunft zu geben. Dass das Zwei-Phasen-Pretesting in dieser Hinsicht erfolgreich arbeitet, hat sich bereits im Rahmen mehrerer Studien erwiesen.

Einschränkend möchten die Autoren allerdings nochmals darauf hinweisen, dass dieses Verfahren aufgrund einer speziellen Konstellation von äußeren Bedingungen entwickelt wurde und deshalb nur eine Kompromisslösung darstellt. Unsere bisherigen Erfahrungen zeigen, dass das Zwei-Phasen-Pretesting zwar durchaus in der Lage ist, Informationen über die zentralen Qualitätsaspekte einer Frage bzw. eines Fragebogens zu liefern, die Verlässlichkeit dieser Informationen darüber, ob die von einer Frage generierten Daten tatsächlich valide und reliabel sind, jedoch oft unbefriedigend ist. Wir meinen deshalb, dass – wenn irgend möglich - größere Anstrengungen unternommen werden sollten, um ein befriedigendes Zuverlässigkeitsniveau der Pretestinformationen zu erreichen. Für den Fall, dass in einem Umfrageprojekt andere, d.h. günstigere Bedingungen herrschen (z.B. größere zeitliche und finanzielle Ressourcen), plädieren wir deshalb dafür, in Phase 1 stärker mit kognitiven Verfahren zu arbeiten (z.B. zusätzlich iterative Vorgehensweise mit Experteninterviews im Labor) und in Phase 2 zusätzliche Techniken wie z.B. Behaviour Coding, Respondent Debriefing oder Interviewer Debriefing einzusetzen.

Pretesting lohnt sich!

Literatur

- Belson, W. A., 1981: The Design and Understanding of Survey Questions. Aldershot, England: Gower.
- Belson, W. A., 1986: Validity in Survey Research. Aldershot, England: Gower.
- Fowler, F. J., 1992: How unclear Terms Affect Survey Data. Public Opinion Quarterly 56: 218-231.
- Fowler, F. J., 1995: Improving Survey Questions. Thousand Oaks: Sage.
- Forsyth, B. H./Lessler, J. T., 1991: Cognitive Laboratory Methods: A Taxonomy. S. 393-418 in: Biemer, P.N. et al. (Hrsg.), Measurement Errors in Surveys. New York: John Wiley.
- Friedrichs, J., 1973: Methoden der Empirischen Sozialforschung. Reinbeck: Rowohlt.
- Hippler, H. J./ Schwarz, N./Sudman, S., 1987: Social Information Processing and Survey Methodology. New York: Springer.
- Jabine, Th. B. /Straf, M. L./Tanur, J. M./Tourangeau, R., 1984: Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. Washington: National Academy Press.
- Jobe, J. B./Loftus, E. F., (Hrsg.), 1991: Cognition and Survey Measurement. Special Issue of Applied Cognitive Psychology 5, 3.
- Karmasin, F./Karmasin, H., 1977: Einführung in die Methode und Probleme der Umfrageforschung. Wien: Hermann Böhlau Nachf.
- Kreiselmaier, J./Prüfer, P./Rexroth, M., 1989: Der Interviewer im Pretest. Mannheim: ZUMA-Arbeitsbericht 89/14.
- Lessler, J.T./Forsyth, B.H., 1995: A Coding System for Appraising Questionnaires. In: N.Schwarz/S. Sudman (Hrsg.), Answering Questions. San Francisco: Jossey-Bass.
- Mohler, P. Ph./Lüttinger, P. (Hrsg.), 2000: Querschnitt. Eine Festschrift für Max Kaase. Neustadt: Pfälzische Post.
- Morton-Williams, J., 1979: The Use of "Verbal Interaction Coding" for Evaluating a Questionnaire. Quality and Quantity 13: 59-75.
- Payne, S. L., 1951: The Art of Asking Questions. Princeton, N.J.: Princeton University Press.
- Presser, S./Blair, J., 1994: Survey Pretesting: Do different Methods produce different Results? Sociological Methodology 1994: 73-104.
- Prüfer, P./Rexroth, M., 1985: Zur Anwendung der Interaction-Coding-Technik. ZUMA-Nachrichten 17: 2-49.

Prüfer, P./Rexroth, M., 1996: Verfahren zur Evaluation von Survey-Fragen: Ein Überblick. ZUMA-Nachrichten 39: 95-116.

Prüfer, P., 1998: Anleitung zur Durchführung standardisierter Interviews. ZUMA-Technischer Bericht 98/02.

Schuman, H., 1966: The Random Probe: A Technique for Evaluating the Validity of Closed Questions. American Sociological Review 31: 218-222.

Schwarz, N./Sudman, S. (Hrsg.), 1996: Answering Questions. Methodology for Determining Cognitive and Communicative Processes in Survey Research. San-Francisco: Jossey-Bass Publishers.

Sudman, S./Bradburn, N., 1982: Asking Questions. A Practical Guide to Questionnaire Design. San Francisco: Jossey-Bass.

Sudman, S./Bradburn, N./Schwarz, N., 1996: Thinking about Answers. The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass.